Proc. Natl. Acad. Sci. USA
Vol. 85, pp. 46–50, January 1988
Biochemistry

Isolation and structural characterization of a cDNA clone encoding rat gastric intrinsic factor

(cobalamin)

B. K. DIECKGRAEFE, B. SEETHARAM, L. BANASZAK, J. F. LEYKAM, AND D. H. ALPERS

Department of Medicine, Washington University School of Medicine, St. Louis, MO 63110

Communicated by David M. Kipnis, September 14, 1987

ABSTRACT **Rat intrinsic factor (IF) has been purified and proteolytic fragments were sequenced. A cDNA library was constructed from size-enriched gastric poly(A)$^+$ RNA and screened for IF-positive clones by antibody and synthetic oligodeoxynucleotide probe hybridization. An IF clone was isolated and sequenced, revealing a predicted primary amino acid sequence in the coding region of 421 amino acids and a putative signal sequence of 22 amino acids. The primary translation product of IF produced in a cell-free translation system displayed cobalamin (Cbl)-binding activity without proteolytic processing or glycosylation. The amino-terminal region of IF showed significant secondary structural and hydropathic homologies with the nucleotide-binding domain in NAD-dependent oxidoreductases. Alignment of the first 80 residues of IF, following the signal peptide, demonstrated homology with the nucleotide-binding domain of cytoplasmic malate dehydrogenase. Based on these data, we propose a model of IF tertiary structure in which the Cbl-binding domain resides in the NH_2-terminal half of the protein.**

Intrinsic factor (IF) is a glycoprotein ranging from 40 to 50 kDa in different mammalian species. It is synthesized in parietal cells in humans, rabbits, and guinea pigs; chief cells in rats and mice; and pyloroduodenal mucus neck cells in hogs (1). The function of IF is to promote absorption of the essential vitamin cobalamin (Cbl) in the ileum by specific receptor-mediated endocytosis (2). IF structure has been preserved among many species, as antiserum against human IF recognizes IF in the hog, dog, and guinea pig gastric mucosa (3). However, despite the importance of this protein, very little is known about its structure and, in particular, its domains for binding Cbl and the ileal receptor. Such an understanding is needed because of the heterogeneous abnormalities in IF that have been reported in patients with Cbl malabsorption (4).

To identify such domains, we have isolated a cDNA clone encoding the entire structural region of rat IF and deduced its primary structure.* We demonstrate that rat IF shows considerable homology with a human IF sequence published earlier, and we propose a model in which a possible binding domain for the nucleotide portion of Cbl is identified.

METHODS

Protein Purification and Peptide Sequencing. Rat IF was purified according to Seetharam *et al.* (5). Peptides were produced by hydrolysis of 2 nmol of IF with endoproteinase Lys-C (Boehringer Mannheim) for 16 hr. Peptides were isolated on sequential anion exchange (TSK-SP5PW) and C-18 reverse-phase (Vydac 218-TP54) HPLC columns (6) and sequenced in an Applied Biosystems (Foster City, CA) 470A automated vapor-phase sequencer (7). Polyclonal antiserum against rat IF was prepared as described (8).

mRNA Isolation and Analysis. Total cellular RNA was extracted by the guanidinium thiocyanate method (9), and poly(A)$^+$ RNA was isolated by binding to oligo(dT)-cellulose (10). RNA was fractionated by electrophoresis in a 1.5% formaldehyde/agarose gel, transferred to a nitrocellulose filter, and hybridized with oligonucleotide or cDNA probes (11). Total gastric RNA was translated in a nuclease-treated reticulocyte lysate system (12) in the presence of 40 units of placental ribonuclease inhibitor per ml. The translation products were isolated either by addition of anti-rat IF and formalin-fixed *Staphylococcus* A cells (Bethesda Research Laboratories) (13), or of Cbl-Sepharose (5) for 1 hr in 100 mM Tris·HCl (pH 3.0) with 100 mM NaCl and 100 μg of bovine serum albumin per ml. Isolated proteins were separated on a 9% $NaDodSO_4$/polyacrylamide gel, fluorographed with En-^{3}Hance (DuPont) and developed at −70°C for 24 hr.

cDNA Library Construction and Screening. Poly(A)$^+$ RNA from rat stomach was enriched for IF mRNA [1.6–2.1 kilobases (kb)] by fractionation, transfer to, and elution from a DEAE membrane, NA45 (Schleicher & Schuell), according to the manufacturer's protocol (ref. 14; Schleicher & Schuell Applications Update 375). Double-stranded cDNA was synthesized using the Amersham cDNA synthesis kit (Amersham cDNA Synthesis System Handbook). Internal *Eco*RI restriction sites were protected by methylation with 40 units of *Eco*RI methylase (New England Biolabs). The cDNA was ligated to kinase-treated *Eco*RI linkers (New England Biolabs) by T4 DNA ligase (11). After removal of linker oligomers, double-stranded cDNA was size selected for molecules >0.7 kb using preparative electrophoresis in low melting point agarose and recovered with an Elutip-d column (Schleicher & Schuell Elutip-d Tech. Bull. 206). The size-selected cDNA was ligated to λgt11 arms (Stratagene, San Diego, CA) and packaged as described (11); 400,000 nonamplified recombinant plaques were screened with anti-rat IF antibody using the Protoblot immunoscreening procedure (Promega Biotec, Madison, WI) (15). Verification of clone identity was obtained by positive hybridization on dot blots probed with a "best-guess" synthetic 35-mer oligonucleotide probe shown in Fig. 1, designed by using a rat codon utilization table and additional theoretical considerations (16). The 35-mer was synthesized in an Applied Biosystems model 380A DNA synthesizer and labeled at the 5′ end with [γ-^{32}P]ATP to a specific activity >1 × 10^6 dpm/pmol (17).

DNA Sequencing and Analysis. The IF-specific clones were sequenced utilizing the dideoxy chain-termination method of Sanger *et al.* (18) in the vector PBSM13+ (Stratagene). Homology searches were carried out against the National Biomedical Research Foundation Protein Sequencing Data-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: IF, intrinsic factor; Cbl, cobalamin.

*This sequence is being deposited in the EMBL/GenBank data base (Bolt, Beranek, and Newman Laboratories, Cambridge, MA, and Eur. Mol. Biol. Lab., Heidelberg) (accession no. J03577).

base, and against the Bolt, Beranek and Newman Genetic Sequence Data Bank (19). Secondary structure predictions were calculated using the method of Chou and Fasman (20). Hydropathy analysis of the IF sequence was done by the method of Kyte and Doolittle (21). Sequence comparison with the nucleotide binding domains in NAD-dependent oxidoreductases was done by using the National Biomedical Research Foundation Protein Identification Resource Alignment Score program version 2.1.

RESULTS

Protein Purification and *in Vitro* Translation of IF. Rat IF (6 mg) was purified from rat stomachs by Cbl-Sepharose affinity chromatography. An attempt was then made to sequence 2 nmol of the intact IF molecule, but no phenylthiohydantoin-derivatized amino acids were detected in the first five cycles analyzed, indicating a blocked NH_2 terminus. Hydrolysis of the sample on the disk verified the presence of the 2 nmol of IF. Internal sequences were obtained using peptides from an endoproteinase Lys-C digestion. Fig. 1 shows one of the peptides and the resulting best-guess oligonucleotide probe containing 35 bases.

Total gastric RNA was translated in a reticulocyte lysate system to demonstrate the Cbl-binding ability of the primary translation product of IF. Fig. 2 shows the ability of the unprocessed IF to bind to Cbl-Sepharose. This binding was stable in 1 M guanidine hydrochloride, but IF was displaced by the addition of free Cbl (lane 2). Although the data are not shown, this same protein could be immunoprecipitated by anti-intrinsic factor antibody and *Staphylococcus* A cells.

The best-guess oligonucleotide probe detected a 1.7-kb sequence by RNA blot analysis (Fig. 3, lane A). This sequence was found only in stomach mRNA and presumably encoded IF. Moreover, *in vitro* translation of fractionated mRNA revealed that IF was encoded by an mRNA species between 1.6 and 2.1 kb long (data not shown).

IF cDNA. Poly(A)$^+$ RNA from 1.6 to 2.1 kb long was recovered from the DEAE membrane after transfer from the agarose gel and was used to construct a library of 1.4×10^6 recombinant clones. Clones (4×10^5) were screened by antibody detection. A total of 8 positive clones were detected after tertiary screening and DNA was purified from each. Dot blot hybridization revealed that they represented two independent groups (4 clones each) that did not cross-hybridize to each other. Members of only one of the two groups were uniformly positive with the 35-mer oligonucleotide probe. cDNA inserts from the oligo-positive antibody-positive group were found to have overlapping segments by restriction analysis and ranged from 0.7 to 1.5 kb long. The size of the stomach mRNA encoding IF was 1.7 ± 0.1 kDa when probed with the full-length clone (Fig. 3, lane D). No hybridizing mRNA species was present in the liver (lane C).

The cDNA inserts of the four positive clones were subcloned and sequenced using double-stranded DNA sequencing methodology. The coding region of the full-length clone

```
Rat IF    Phe Gln Asn Pro Met Ser Ile Ala Gln Ile Leu Pro Ser Leu Lys

mRNA  5' UUC CAG AAC CCC AUG UCC AUC GCC CAG AUC CUG CC  3'
           U   A   U   U     AGU   U   U   A   U U U
               A       A   A   A       A   C
               G       G       G           A

Oligo 3' AAG GTC TTG GGG TAC AGG TAG CGG GTC TAG GAC GG  5'

cDNA  5' TTC CAG AAC CCC ATG TCC ATT GCC CAG ATT CTC CC  3'
```

FIG. 1. Best-guess oligonucleotide probe used to verify the anti-intrinsic factor antibody-selected clones. Codon choice in ambiguous positions was dictated by a codon utilization frequency table constructed from 3132 residues of rat proteins taken from the GenBank data base. Comparison with the isolated cDNA is shown. Mismatches are indicated by underlining.

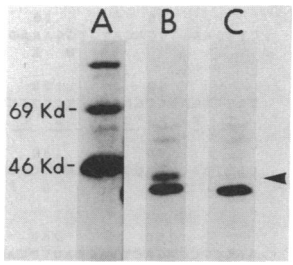

FIG. 2. Isolation of rat IF from a cell-free translation mixture by Cbl-Sepharose. IF synthesis was directed as described in *Methods*. Protein standards (in kDa) are shown in lane A. Lane B, protein isolated by the addition of Cbl-Sepharose beads to the translation mixture, followed by extensive washing, and elution of the bound products by boiling in 0.1% $NaDodSO_4$; lane C, same as lane B, with the addition of 10 μg of cold Cbl added during the incubation with cbl-Sepharose beads. The band corresponding to IF is shown with an arrowhead. This is the same peptide that is precipitated using antiserum to IF. The other major band present has been identified by hybrid-arrested translation as pepsinogen (data not shown).

pIFQ was sequenced in both directions utilizing subcloned fragments and oligonucleotide probes (Fig. 4). The 3′ untranslated region consists of nearly 200 nucleotides up to the poly(A) tail and contains two AATAAA polyadenylylation signals 14 and 60 nucleotides upstream from the start of the poly(A) tail. The deduced unprocessed protein contains 421 amino acids and has a calculated molecular mass of 46,458 Da. This is in good agreement with the primary translation product, which has an apparent molecular mass of ≈44 kDa as determined by $NaDodSO_4$/PAGE (Fig. 2).

The predicted amino acid sequence includes the sequences of two peptides obtained from rat IF and regions of human IF homologous to the NH_2 terminus and four internal fragments (22). An open reading frame begins at the initial ATG starting at nucleotide 13. The sequence upstream from this does not conform to the consensus sequence CC(A/G)CCAUG proposed for eukaryotic initiation sites by Kozak (24). Ten percent of mRNAs examined had a purine in position −3 but otherwise differed entirely from this −1 to −5 consensus sequence. IF mRNA may fit in this group. Another methionine four amino acids downstream, which is not preceded by any components of the consensus sequence, seems unlikely to be the initiation codon, since translation begins at the 5′

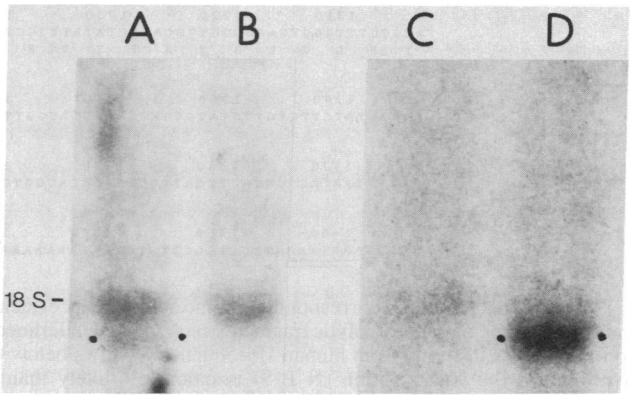

FIG. 3. RNA blot analysis of total RNA. Only the portion of the gel near the 18S ribosomal subunit is shown. Lanes A and B, hybridization to stomach and submaxillary total RNA using the 35-mer oligonucleotide probe. Lanes C and D, hybridization to liver and stomach RNA using the full-length IF clone as a probe. Dots are used to indicate the position of the IF mRNA species. The presence of the 18S rRNA bands in both lanes A and B is due to the low-stringency hybridization and washing conditions necessitated by the use of a best-fit oligonucleotide probe.

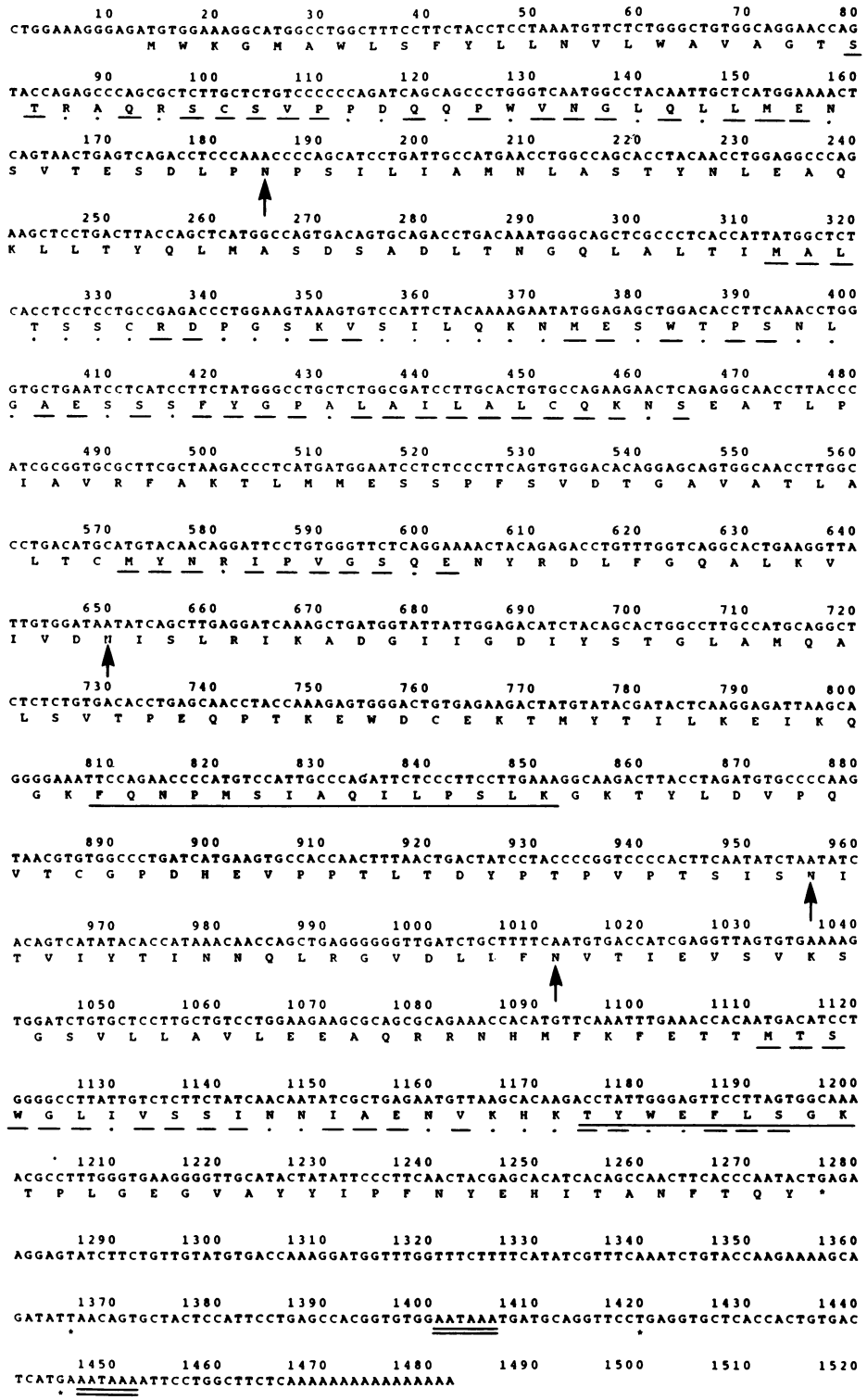

FIG. 4. Nucleotide and corresponding amino acid sequence (identified by the single-letter code) of the full-length clone pIFQ. Solid underlines indicate location of the proteolytic fragments sequenced. Underlined individual residues indicate homology with human IF (22). Dots correspond to the residues differing from human IF sequence. Arrows below the Asn residues indicate possible N-glycosylation sites. The first site with a proline in the Xaa position (N P S) is much less likely than the other three, but at least one such example has been reported (23). Polyadenylylation signals are indicated by solid double underlining.

proximal AUG triplet in ≈95% of the mRNAs tabulated (24). An upstream methionine, not present in this clone, acting as the initiator seems very unlikely, as the four codons upstream would include at least two glutamic acid residues (5′ Leu-Glu-Arg-Glu-), which would not fit into a putative NH_2-terminal signal peptide.

Amino Acid Sequence of IF. The NH_2 terminus of the mature human IF begins with the Ser-Thr found at amino acid residue 23 in the rat, corresponding with nucleotide residues 79–84 (Fig. 4). The sequence toward the NH_2-terminal end of this region is characteristic of a signal sequence, with a basic NH_2-terminal region (Lys) at nucleotides 19–21, a central hydrophobic domain (Ala, Leu, Val), and a more polar COOH-terminal region (Gly, Thr, Ser) at nucleotide residues

73–81. Based on statistics compiled by von Heijne (25) for an acceptable signal-sequence cleavage site, the side chain of the residue in position −1 must be small, and the residue at −3 is usually Gly or Ala. Cleavage of Thr-Ser conforms to this requirement. However, the results of a weighted matrix analysis of possible signal peptide cleavage sites revealed that the most likely site of cleavage is predicted between Thr-24 and Arg-25 (nucleotide residues 82–84), with a score of 6.7. The sites of cleavage that would generate the human IF NH_2 terminus, Thr-22/Ser-23 and Ala-18/Val-19, were also strongly predicted with scores of 4.4 and 4.1, respectively. The sequence of IF has four possible N-linked glycosylation sites, as shown in Fig. 4, corresponding to the consensus sequence of Asn-Xaa-Thr or Asn-Xaa-Ser. The first of these (Asn-Pro-Ser) contains a proline, an unlikely residue for Xaa, but an apparent site (Asn-Pro-Thr) for partial glycosylation in horse ribonuclease (23).

Secondary Structure Prediction and Comparison with Other Proteins. The National Biomedical Research Foundation protein sequence data base and the Bolt, Beranek, and Newman genetic sequence data base were searched using the full-length deduced amino acid and nucleotide sequence of IF. No striking amino acid sequence homologies were detected. Matrix analysis was used to compare the sequence against itself. No significant repeating domains were found. Since Cbl appeared to have similarities to NAD^+, the coenzyme for dehydrogenases, a more detailed comparison of IF was made with that of malate dehydrogenase.

The first 80 residues of intrinsic factor could be aligned with the nucleotide-binding region of cytoplasmic malate dehydrogenase, using the alignment score program from the Protein Identification Resource (National Biomedical Research Foundation) and a mutational data matrix (26) (Fig. 5). With a gap penalty of 6 (a comparison parameter that decrements the matching score to control for amino acid substitutions), 26% of the aligned residues were homologous, and >80% of the aligned residues were either identical or could result from a single base-pair change in the codon. Comparison of the two proteins beyond the NH_2-terminal domain revealed no significant additional amino acid homology. A hydropathy index was calculated by the method of Kyte and Doolittle (21) with a 5-residue window. Significant hydrophobicity, defined as a score of >1.0, is indicated for both proteins in the same figure and these appear to correlate well. Furthermore, the Chou–Fasman secondary structure predictions for β-strands are also shown for IF along with the actual location of β-strands and α-helices for cytoplasmic malate dehydrogenase.

A comparison of the benzimidazole-ribofuranosyl portion of cobalamin was made with the adenine ribose of NAD. This analysis was carried out using Sybyl version 3.4 (Tripos Associates, St. Louis, MO). Coordinates of the benzimidazole-ribofuranosyl moiety were taken from the crystal structure of cyano-13-epicobalamin and coordinates of NAD^+ as bound to crystalline malate dehydrogenase (27). After visual alignment of the two molecules, the calculated root-mean-square distance between corresponding atoms was 1.05 Å.

DISCUSSION

The isolation of the cDNA clone encoding IF, represented by a nonabundant mRNA, was made easier by the correct identification of the mRNA size (Fig. 3). Although best-guess oligonucleotides have been often used to screen complete cDNA libraries, in this instance the oligonucleotide served a dual function, allowing first the size fractionation of mRNA as a source of the library. The RNA blot reading was validated by the correct identification of the IF clone by the same oligonucleotide (Fig. 3, lane C). This use of oligonucleotides may prove helpful in identifying cDNA clones corresponding to other rare mRNA species.

One of the plasmid clones, pIFQ, had a 1.5-kb insert containing the entire coding region for intrinsic factor, as well as a short poly(A)$^+$ segment and both 3′ and 5′ nontranslated regions. The remarkable resistance of intrinsic factor to proteolysis by pancreatic proteases and pepsin suggested the absence of available cleavage sites for these enzymes. A search of the primary amino acid sequence revealed multiple potential cleavage sites for pepsin, trypsin, and chymotrypsin. This suggests that the resistance to these enzymes is more likely due to tertiary configuration and/or glycosylation.

The presence of the 1-α-D-ribofuranosyl-5-6-dimethyl benzimidazole-3-phosphate portion of the Cbl molecule prompted us to look for structural similarities between IF and nucleotide-binding proteins. All previously studied structurally known proteins that bind nucleotides demonstrate a common structural theme (28), consisting of a conserved series of parallel-stranded β-strands interconnected by α-helices. Despite identical topology, there is little detectable sequence homology between these nucleotide-binding proteins (29). It is clear from comparisons of homologous proteins that three-dimensional structures are conserved in evolution more than both primary protein sequences and nucleotide sequences (30). The hypothesis that IF has an NH_2-terminal domain conformationally similar to the nucleotide binding proteins is supported by several lines of evidence presented in this paper.

(*i*) The adenine–ribose–pyrophosphate portion of NAD bound in extended conformation in the dehydrogenases is similar in size, charge, and conformation to the benzimidazoyl–ribofuranosyl–phosphate portion of Cbl.

(*ii*) The predicted secondary structure of IF appears to correspond to that of cytoplasmic malate dehydrogenase

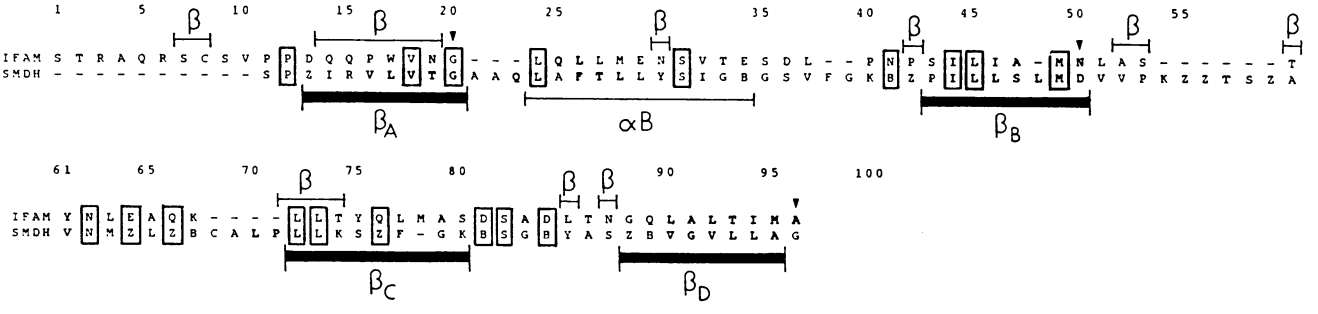

FIG. 5. Alignment of the nucleotide binding domain of cytoplasmic malate dehydrogenase (SMDH) with the first 80 amino acids (identified by the single-letter code) of IF (IFAM) following the signal peptide. Identical conserved amino acids that are present in both proteins are indicated by open boxes. Position of β-strands A–D and the α-helix labeled B in SMDH (27) and in IF [predicted by Chou and Fasman (20)] are shown below and above the sequence comparisons, respectively. Positions of hydrophobic segments for both proteins, calculated by the method of Kyte and Doolittle (21), are indicated by shading. Positions of the conserved residues discussed in the text are indicated by arrowheads.

(Fig. 5). The consensus nucleotide-binding conformation is shown schematically in Fig. 6. This portion includes β-strands β_A, β_B, β_C, and β_D, and the invariant α-helix between strands A and B. In this domain, the α-helix is thought to provide a dipole moment positive near the NH_2 terminus of strand β_B, which interacts with the negatively charged phosphate moiety. A similar function would be required for the binding of Cbl by IF.

(*iii*) In NAD-dependent dehydrogenases, the adenine-binding site is a hydrophobic cavity formed mainly by residues in β strands B and D. As expected, the IF sequence is hydrophobic in these regions and is more similar to cytoplasmic malate dehydrogenase than in the intervening sequences (Fig. 5). Previous work by Kolhouse and Allen (31) has shown that Cbl analogues with diverse alterations in the dimethyl benzimidazole moiety are still bound by IF with high affinity, suggesting that these interactions are relatively nonspecific.

(*iv*) NAD-dependent dehydrogenases have several conserved amino acid residues, all of which are present or conservatively substituted in IF. The first invariant residue is the last residue of the first β-strand of the Rossmann fold, β_A (Fig. 6). In this location, a glycine is necessary to avoid steric interaction with the ribosyl moiety. A glycine is also present in an identical location of IF. Generally, the last residue of the second β-strand, β_B, is also conserved. In this position in the NAD-dependent dehydrogenases an aspartate is present that hydrogen bonds to the O-2′ of the adenine ribose. In the same position, IF contains an asparagine that potentially serves the same function in ligand binding. Finally, glycine is found at the end of β-strand D in the dehydrogenases and in IF, a small residue, alanine, is found in the equivalent position.

Assuming the similarity to the nucleotide-binding proteins is correct, the NH_2-terminal sequence of IF contains the structural requisites necessary for Cbl binding. The homology noted between the Cbl-binding portion of IF and the NAD-binding domain of cytoplasmic malate dehydrogenase suggests that these domains evolved from a common ancestral protein and have analogous functions. Additional studies using directed mutagenesis and analysis of mutant proteins are needed to identify amino acid residues that interact with the corrin ring of Cbl and those that form the receptor-binding domain of IF.

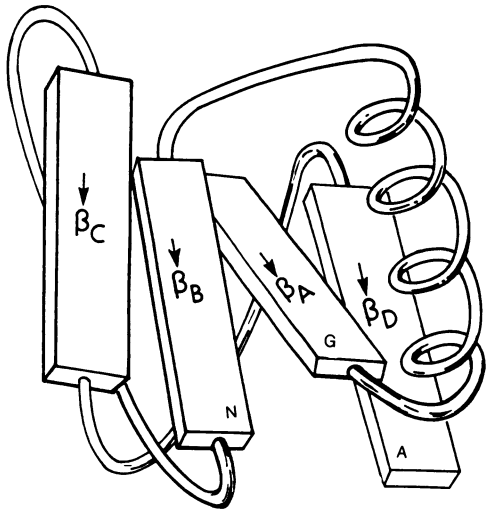

FIG. 6. Tertiary structure model of the NH_2-terminal region of IF. This model was constructed by examining homologies (residue alignment, predicted secondary structure, hydropathy, and functionally conserved residues) with NAD-dependent oxidoreductases. Functionally conserved residues are indicated. Adapted from Birktoft and Banaszak (29).

We thank Joyce Rudolph for manuscript preparation and Jim Dunbar for assistance with the Sybyl comparison. This work was supported in part by Grants DK14038 and DK33487 from the National Institutes of Health, and by the Monsanto Company (St. Louis, MO). We acknowledge the support of the National Science Foundation (DMB-828894) for related studies in dehydrogenases.

1. Hoedemaker, P. J., Abels, J., Wachters, J. J., Arends, A. & Nieweg, H. O. (1966) *Lab. Invest.* **15,** 1163–1171.
2. Kapadia, C. F., Serfelippi, D., Voloshin, K. & Donaldson, R. M. (1983) *J. Clin. Invest.* **71,** 440–448.
3. Glass, G. B. J. (1984) *Gastric Intrinsic Factor and Other Vitamin B12 Binders* (Georg Thieme, Stuttgart, F.R.G.).
4. Kapadia, C. F. & Donaldson, R. M. (1985) *Annu. Rev. Med.* **36,** 93–110.
5. Seetharam, B., Bakke, J. E. & Alpers, D. H. (1983) *Biochem. Biophys. Res. Commun.* **115,** 238–244.
6. Leykam, J. F. & Schwartz, B. D. (1986) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **45,** 1634.
7. Hewick, R. M., Hunkapiller, M. W., Hood, L. E. & Dreyer, W. J. (1981) *J. Biol. Chem.* **256,** 7990.
8. Seetharam, B., Alpers, D. H. & Allen, R. H. (1981) *J. Biol. Chem.* **256,** 3785–3790.
9. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18,** 5294–5299.
10. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69,** 1408–1412.
11. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY).
12. Pelham, H. R. B. & Jackson, R. J. (1976) *Eur. J. Biochem.* **67,** 247–256.
13. Kessler, S. W. (1975) *J. Immunol.* **115,** 1617–1624.
14. Holland, L. T. & Wangh, L. J. (1983) *Nucleic Acids Res.* **11,** 3283–3300.
15. Young, R. A. & Davis, R. W. (1983) *Science* **222,** 778–782.
16. Lathe, R. (1985) *J. Mol. Biol.* **183,** 1–12.
17. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
18. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
19. Wilbur, W. J. & Lipman, D. J. (1983) *Proc. Natl. Acad. Sci. USA* **80,** 726–730.
20. Chou, P. Y. & Fasman, G. D. (1978) *Annu. Rev. Biochem.* **47,** 251–276.
21. Kyte, J. & Doolittle, R. F. (1982) *J. Mol. Biol.* **157,** 105–132.
22. Nexo, E., Olesen, H., Hansen, M. R., Bucher, D. & Thomsen, J. (1976) *Scand. J. Clin. Lab. Invest.* **38,** 649–653.
23. Scheffer, A. J. & Beintema, J. J. (1974) *Eur. J. Biochem.* **46,** 221–233.
24. Kozak, M. (1984) *Nucleic Acids Res.* **12,** 857–872.
25. Von Heijne, G. (1986) *Nucleic Acids Res.* **14,** 4683–4690.
26. Dayhoff, M. O., Schwartz, R. M. & Orcott, B. C. (1978) in *Atlas of Protein Sequence and Structure,* ed. Dayhoff, M. O. (Natl. Biochem. Res. Found., Washington, DC), Vol. 5, Suppl. 3, pp. 345–362.
27. Hill, E., Tsernoglou, D., Webb, L. & Banaszak, L. J. (1972) *J. Mol. Biol.* **72,** 577–591.
28. Rao, S. T. & Rossmann, M. G. (1973) *J. Mol. Biol.* **76,** 241–256.
29. Birktoft, J. J. & Banaszak, L. J. (1984) *Pept. Protein Rev.* **4,** 1–46.
30. Bajaj, M. & Blundell, T. L. (1984) *Annu. Rev. Biophys. Bioeng.* **13,** 453–492.
31. Kolhouse, J. F. & Allen, R. H. (1977) *J. Clin. Invest.* **60,** 1381–1392.